

## 内容感知的可解释性路面病害检测模型

李 傲<sup>1,2</sup> 葛永新<sup>1</sup> 刘慧君<sup>3</sup> 杨春华<sup>3</sup> 周修庄<sup>4</sup>

<sup>1</sup>(重庆大学大数据与软件学院 重庆 401331)

<sup>2</sup>(重庆城市管理职业学院智能工程学院 重庆 401331)

<sup>3</sup>(重庆大学计算机学院 重庆 400044)

<sup>4</sup>(北京邮电大学人工智能学院 北京 100876)

(li\_ao@cqu.edu.cn)

## Content-Aware Explainable Pavement Distress Detection Model

Li Ao<sup>1,2</sup>, Ge Yongxin<sup>1</sup>, Liu Huijun<sup>3</sup>, Yang Chunhua<sup>3</sup>, and Zhou Xiuzhuang<sup>4</sup>

<sup>1</sup>(School of Big Data & Software Engineering, Chongqing University, Chongqing 401331)

<sup>2</sup>(School of Intelligent Engineering, Chongqing City Management College, Chongqing 401331)

<sup>3</sup>(College of Computer Science, Chongqing University, Chongqing 400044)

<sup>4</sup>(School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876)

**Abstract** To address the challenges of using high-resolution pavement images as input for existing convolutional neural network models and the inability of existing preprocessing algorithms to effectively perceive and retain information from low-ratio distress regions in original pavement images, a novel architectural unit called adaptive perception module (APM) paying greater attention to pavement distress region is proposed with the help of visual interpretation techniques, which achieves a rapid and accurate detection of pavement distress in high-resolution images and could be used to build a software system for automatic detection of pavement distress based on computer vision. Firstly, big kernel convolution and residual operations are used to reduce the origin image resolution and get the low-level but rich feature representation. Secondly, attention mechanism is developed to perceive and activate the region of pavement distress and filter the irrelevant background pixel noise. By means of joint learning, APM training could be completed without additional cost. After the visual interpretation method is used to aid the selection and design of the specific structure of APM, experimental results on the latest public dataset CQU-BPMDD show that the proposed APM significantly improves the classification accuracy, up to 84.47%. Experiments across different datasets CQU-BPDD demonstrate the generalization and robustness of APM. Code is available on <https://github.com/Li-Ao-Git/apm>.

**Key words** pavement distress detection; explainability; adaptive perception; attention mechanism; joint learning

**摘要** 针对实际场景中高分辨路面图像难以直接作为现有卷积神经网络（convolutional neural network, CNN）的输入、现有预处理及下采样算法无法有效感知并保留原始路面图像中低占比的病害区域信息等问题，借助于可视化解释的技术手段，设计了一种即插即用的图像内容自适应感知模块（adaptive

收稿日期：2022-09-06；修回日期：2023-05-15

基金项目：国家自然科学基金项目(62176031, 61972046); 重庆市自然科学基金项目(CSTB2022NSCQ-MSX1405); 重庆市技术创新与应用发展专项项目(CSTB2022TIAD-KPX0100)

This work was supported by the National Natural Science Foundation of China (62176031, 61972046), the Natural Science Foundation of Chongqing (CSTB2022NSCQ-MSX1405), and the Specialized Project for Technology Innovation and Application Development of Chongqing (CSTB2022TIAD-KPX0100).

通信作者：葛永新([yongxinge@cqu.edu.cn](mailto:yongxinge@cqu.edu.cn))

perception module, APM), 既平衡了高分辨路面图像与 CNN 输入限制, 又能够自适应感知激活前景病害区域, 从而实现高分辨路面图像中病害类型的快速准确检测, 构建可信路面病害视觉检测软件系统。APM 利用大卷积核和下采样残差操作降低原始图像分辨率并获取图像浅层特征表示; 通过注意力机制自适应感知并激活图像中路面病害区域信息, 过滤无关的背景信息。利用联合学习的方式, 无需额外监督信息完成对 APM 的训练。通过可视化解释方法辅助选择和设计 APM 的具体结构, 在最新公开数据集 CQU-BPMDD 上的实验结果表明: APM 相比于现有的图像预处理采样算法均有明显提升, 分类准确率最高为 84.47%; 在 CQU-BPDD 上的实验结果及 APM 决策效果可视化分析表明 APM 具备良好的泛化性与鲁棒性。实验代码已开源: <https://github.com/Li-Ao-Git/apm>。

关键词 路面病害检测; 可解释性; 自适应感知; 注意力机制; 联合学习

中图法分类号 TP391

公路交通不仅关系到一个国家和地区的经济命脉, 同时对人民生活水平的提高也发挥着无法替代的作用。我国公路多为沥青路面, 由于其自身空隙大、温度稳定性差、耐衰老性差等缺点, 极易造成裂纹、疏松等病害<sup>[1]</sup>, 这些路面病害严重影响了公路的承载力、运输力、客货运的行车安全以及公路使用年限。因此, 对路面病害快速准确的有效检测是公路运输安全的重要保障。

传统人工检测方法严重依赖检测人员的主观判断, 存在漏检错检率高、效率低、检测成本高昂等问题, 难以满足精确、高效、低成本的路面多类别病害检测需求<sup>[2-3]</sup>。

随着计算机视觉技术的不断发展, 利用其对路面病害进行自动化检测在技术上逐渐成熟可行。借助计算机视觉技术实现路面病害自动化检测软件系统的流程遵循一般的软件工程开发流程<sup>[4]</sup>。本文的主要工作集中在路面病害检测算法的设计与验证。可信路面病害自动化检测软件系统中对于算法设计与验证部分提出了明确的需求, 即构建具备良好泛化性和鲁棒性的路面病害多分类检测模型。

基于计算机视觉技术的路面病害检测算法在流程上可分为特征提取、阈值分类 2 部分, 依据特征提取模式和分类阈值设定方式上的不同可以分为传统方法、基于机器学习的方法和基于深度学习的方法。传统方法主要依靠人工提取图像特征并设定分类阈值, 比如 Amhaz 等人<sup>[5]</sup> 提取病害图像的光照强度和几何学特征用于病害分类, 受限于人工提取图像特征和设定分类阈值的方式, 传统方法无法有效处理复杂环境路况下的病害分类任务, 导致在大规模数据下的检测准确率偏低。Koch 等人<sup>[6]</sup> 在路面病害检测任务中利用机器学习方法自动设定分类阈值; 后续研究者 Kapela 等人<sup>[7]</sup> 则使用支持向量机得到分类

阈值; 也有研究者尝试改进贝叶斯分类器<sup>[8]</sup> 或 AdaBoost 分类器<sup>[9]</sup> 获取分类阈值。基于机器学习的方法虽然通过学习来获取设定分类阈值, 但依旧无法摆脱人工提取图像特征的困境, 难以实现复杂路况下路面病害的高效检测。而深度学习技术借助卷积神经网络(convolutional neural network, CNN)能够学习到丰富、多层次的图像特征表示, 并通过全卷积或全连接的方式对图像特征进行自动分类检测, 整个过程无需人工介入, 因此越来越多地被用于路面病害检测。Gopalakrishnan 等人<sup>[10]</sup> 利用预训练的 VGG-16 网络可以高效地检测出路面是否含有裂缝; Li 等人<sup>[11]</sup> 和 Cha 等人<sup>[12]</sup> 对主流网络模型进行修改和微调也能够准确地检测出路面是否包含病害; Tang 等人<sup>[13]</sup> 提出一种迭代优化的网络模型, 进一步提高了路面病害二分类的准确率。利用模型自动学习来提取图像特征和设置分类阈值的深度学习技术成为主流的研究算法, 其对路面病害的分类准确率远高于传统方法和机器学习。

目前主流的 CNN 网络多采用  $224 \times 224$  或  $256 \times 256$  等低分辨率图像作为输入, 实际应用场景中为提高图像采集效率和图像质量, 所采集的路面图像分辨率往往为  $3692 \times 2147$  甚至更高, 无法将实际场景中采集到的路面图像直接输入到现有的预训练网络中。对此, 目前主要有 3 种解决办法。第 1 种处理方式是修改现有网络并重新训练, 使得原始图像可以直接输入到修改后的网络中, 但这种方式会引入大量无意义计算, 提高网络训练成本、降低推理速率, 鲜有研究者尝试这种处理方式。第 2 种处理办法是将原始图像分块, 例如 Tang 等人<sup>[13]</sup> 和 Huang 等人<sup>[14]</sup> 将原始图像进行分块, 使用分块后的图像训练网络。但是这种分块方法存在 3 个问题: 1) 图像的每一块缺乏用于网络训练的监督标签; 2) 每一块独立输入网络未

考虑到各块之间的联系，丢失整体结构信息；3)仍将图像全部像素输入网络，整体计算量没有降低。其中第3种也是目前主流的处理方法，即对原始图像直接进行下采样。传统的下采样算法如最近邻、双线性插值等，依据采样前后像素点数目的比值计算像素信息丢失率，通常会损失超过95%的像素信息，前景病害信息和无关背景信息均同等程度受损，且受噪

声与背景变化影响较大。Talebi等人<sup>[15]</sup>为了尽可能保留原始图像中所有细节信息，提出了一种可学习的下采样器，其采样率在0.25~1之间。但病害像素在原始图像中的分布呈连续集中状态且占比低，如图1所示，无选择地保留全部细节信息意味着同等对待无关背景区域，缺乏对前景病害区域的关注，无法有效解决路面病害检测任务中的痛点问题。

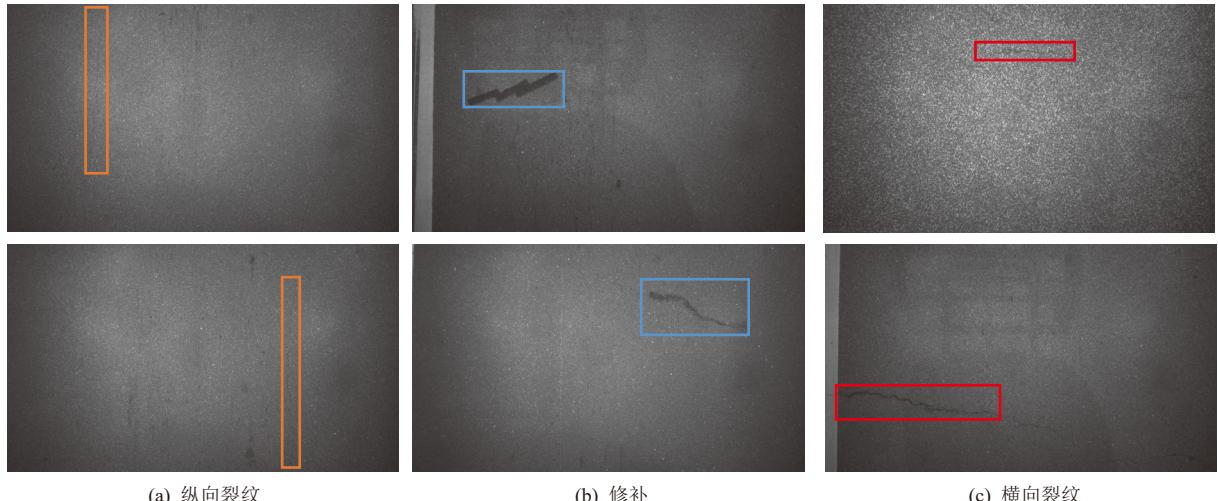


Fig. 1 Pavement distress images in CQU-BPMDD

图1 CQU-BPMDD 中的路面病害图像

快速有效地设计性能高效的网络模型既可以平衡高分辨率图像与CNN的输入限制，又能自适应感知并保留病害区域信息。可视化解释方法通过对CNN的工作原理以及决策过程的解释性分析<sup>[16]</sup>，更好地辅助选择和设计网络模型。以往的研究中可视化解释方法通常是作为一种事后的解释方法<sup>[13-14]</sup>，用于模型效果的补充验证。可视化解释方法同样可以作为事前的分析方法服务于模型的设计，从而可以更好地构建可信路面病害检测网络模型。

基于以上研究，本文借助于可视化解释方法设计了一种即插即用的图像内容自适应感知模块(adaptive perception module, APM)来替换现有预处理算法，该模块自适应感知并保留病害区域相关信息，显著提升高分辨图像下的路面病害分类准确率。本文的主要贡献有3个方面：

1) 借助于可视化解释方法设计了图像内容APM。APM借助残差模块获取更为丰富的特征表示，利用卷积注意力模块(convolutional block attention module, CBAM)<sup>[17]</sup>来感知激活病害区域的特征信息，过滤无效的背景信息，获取低分辨率下包含丰富病害区域信息的图像数据。

2) 以联合学习的方式对APM进行训练，APM可以即插即用地嵌入到现有骨干网络(Backbone)之前。无需额外的监督信息与设计独立的损失函数，即可实现端到端的路面病害检测模型训练，从而降低整体的训练成本。

3) 在最新公开的数据集CQU-BPMDD<sup>[18]</sup>上的实验表明本文提出的APM相比于现有预处理算法能够有效地保留病害区域的特征信息，路面病害检测准确率提升明显。在数据集CQU-BPDD<sup>[13]</sup>上的实验及可视化分析表明APM具备良好的泛化性与鲁棒性。

## 1 相关工作

### 1.1 基于CNN的路面病害检测

目前已有较多的工作将卷积神经网络等深度学习技术应用到路面病害检测任务，相较传统模型取得了较大提升<sup>[19]</sup>。依据数据集提供的标注信息的粒度不同，可以将这些方法分为2类：像素级别的分类检测和图像级别的分类检测。

像素级别的分类检测需要为路面图像中的每一个像素点分配类别标签，也称为语义分割。目前主流

的研究方法主要将其他领域的语义分割模型迁移到当前任务中。Jenkins 等人<sup>[20]</sup>迁移 U-Net 模型实现病害区域分割。Yang 等人<sup>[21]</sup>利用融合特征金字塔模型的 CNN 架构来实现病害区域的分割。Liu 等人<sup>[22]</sup>和 Zhang 等人<sup>[23]</sup>在编解码结构中加入注意力机制分别提出了 FPCNet 和 DeepCrack, 同时, Xiang 等人<sup>[24]</sup>在金字塔特征结构的基础上融合注意力机制进行病害区域分割。此外, 全卷积网络也常常被用于像素分割<sup>[25]</sup>。但像素级别分类需要准确的像素级别的标注信息来训练, 实际场景中往往缺乏像素级别的标注信息。

图像级别的分类检测主要判断图像中是否包含路面病害。Zhang 等人<sup>[26]</sup>和 Tang 等人<sup>[13]</sup>利用图像分块的思想, 将不同的原始图像分割成多个重叠子区域, 通过对子区域的检测来判断原始图像是否包含路面病害。Krizhevsky 等人<sup>[27]</sup>自行设计 CNN 模型探索网络深度和分类准确率之间的关系。主流的研究主要集中在对路面有无病害的二分类检测, 由于缺乏多分类数据集, 目前对路面病害多分类检测的研究较少。

## 1.2 图像缩放技术

主流的 CNN 网络多采用  $224 \times 224$  或  $256 \times 256$  等低分辨率图像作为输入, 但实际场景中采集到的图像分辨率取决于具体的采集设备, 并不一定严格满足 CNN 的输入限制条件, 这种情况下需要利用图像缩放技术来对原始图像进行缩放。

传统图像缩放算法如最近邻、双线性或双三次插值技术, 依据一定的计算规则来得到目标图像的像素值, 这些算法不需要监督信息, 具有计算量小、计算速度快等优势, 使其成为深度学习中图像预处理阶段最为常用的缩放办法。但传统的缩放算法并没有考虑到图像中不同目标内容在缩放过程中的重要程度, 无法依据具体的图像内容数据进行变化, 仅仅是无差别处理无关背景及病害区域前景, 在路面病害自动化检测的软件系统中应用有限。

Talebi 等人<sup>[18]</sup>基于 CNN 提出了一种可学习的下采样算法, 在缩放过程中关注图像的整体细节信息, 保留图像更多的细节信息从而优化网络性能。CNN 虽然可以通过学习获取图像特征语义表示, 但是并未考虑感知图像中与具体任务紧密相关的信息。

## 1.3 可视化解释方法

CNN 模型的可解释性问题, 也称为深度可视化问题。依据计算过程, 可以将深度可视化方法分为 2 大类。一类通过正向计算直接可视化 CNN 网络每一层的卷积核参数以及输出的特征图; 另一类通过反

向计算, 将高维特征反向到 CNN 的每一层卷积, 可以观察特征图对应原图的区域, 从而了解 CNN 从原图中学习到的具体特征信息。

最早的深度可视化工作通过正向过程可视化卷积核参数<sup>[28]</sup>, CNN 的浅层卷积主要关注于图像中边缘、条纹以及颜色信息。Zeiler 等人<sup>[29]</sup>则是对正向过程中各卷积层输出的特征图进行可视化, 分析不同通道对于输入图像的差异性响应。Zeiler 等人<sup>[30]</sup>同时对卷积核参数与特征图进行可视化分析, 揭示不同的 CNN 网络结构激发映射并且揭示出其对输入变形的不变性。这种通过正向计算的可视化不仅需要分析大量的卷积核参数, 同时必须选择充足的样本进行特征提取, 而 CNN 网络中存在很多冗余, 导致可视化中存在较多的无效工作。

另一类可视化借助于反向计算, 经典方法有反卷积<sup>[31]</sup>和梯度反向传播。Zeiler 等人<sup>[32]</sup>通过反卷积来对图像特征进行重建, 分析输入图像中的哪些信息被保留在 CNN 提取的特征中。Zhou 等人<sup>[33]</sup>提出类激活映射(class activation mapping, CAM)可视化方法, 使用全卷积层替换神经网络末端的全连接层, 并将输出层的权重反向投影至卷积层特征, 能有效定位图像中有助于分类任务的关键区域, 但是 CAM 需要对模型进行修改。Grad-CAM<sup>[34]</sup>应运而生, 用梯度的全局平均来计算特征图对应的权重, 最后加权求和并叠加至原始输入, 有效定位图像中有助于分类任务的关键区域, 无需修改原始模型也无需重新训练网络, 成为目前深度可视化问题中主流的选择。本文选择 Grad-CAM 作为可视化解释的技术手段, 辅助 APM 模块的设计及效果验证。

在路面病害检测任务中, 大多数研究者追求模型在特定测试集上的高准确率, 可视化解释手段作为一种事后的补充证明, 是事后解释性的一种手段。本文则是考虑利用可视化解释方法来理解分析 CNN 的工作原理以及决策过程, 进而更好地选择和设计网络。

## 2 内容感知的可解释性路面病害图像检测模型

本节主要关注内容感知的路面病害图像检测模型的设计, 并借助于可解释性方法构建 APM 模型。

### 2.1 总体框架

本文的总体框架结构如图 2 所示, 整体结构由 APM 模块和分类骨干网络(Backbone)两部分组成,

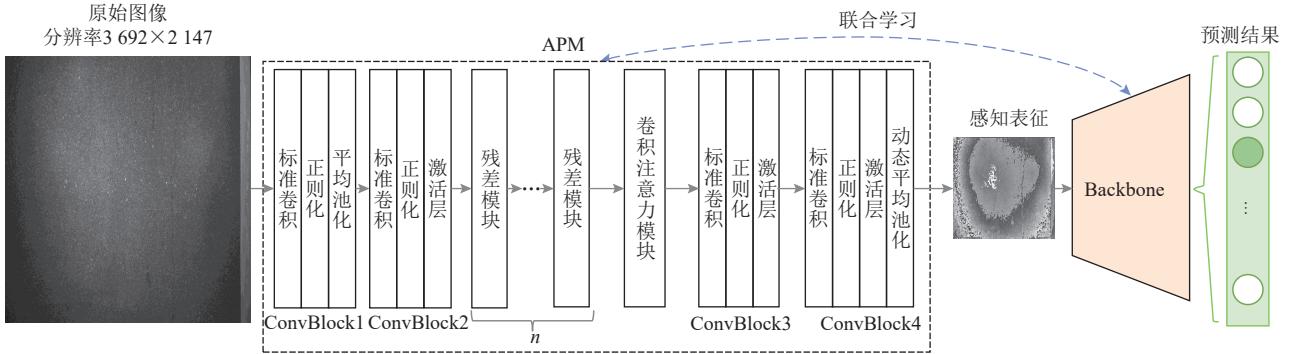


Fig. 2 Overall framework of pavement distress detection model

图2 路面病害检测模型整体框架

APM 和 Backbone 采用联合学习的方式进行训练.

APM 中残差模块的数目可以依据实际场景进行调整, APM 末端添加动态平均池化操作, 将 APM 的输出池化为 Backbone 所需的大小, 实现即插即用地

$$\begin{cases} H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} + 2 \times \text{padding} - \text{dilation} \times (\text{kernel\_size} - 1)}{\text{stride}} + 1 \right\rfloor, \\ W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} + 2 \times \text{padding} - \text{dilation} \times (\text{kernel\_size} - 1)}{\text{stride}} + 1 \right\rfloor. \end{cases} \quad (1)$$

本文通过调整卷积操作的参数来降低输入数据的分辨率. 卷积操作的计算逻辑如式(1)所示, 其中,  $H_{\text{in}}, W_{\text{in}}$  分别表示输入图像的高度和宽度;  $H_{\text{out}}, W_{\text{out}}$  分别表示输出图像的高度和宽度;  $\text{padding}$  表示图像边界填充方式;  $\text{dilation}$  表示卷积膨胀系数;  $\text{kernel\_size}$  表示卷积核大小;  $\text{stride}$  表示卷积步长. 设置  $\text{padding} = 0$ ,  $\text{dilation} = 0$ ,  $\text{stride} = 2$ , 将输出图像的  $H_{\text{out}}, W_{\text{out}}$  分别转换为输入图像  $H_{\text{in}}, W_{\text{in}}$  的  $1/2$ , 从而达到下采样的效果.

注意力机制可以促使模型聚焦于目标区域, 过滤无关背景区域, 增强图像的特征表示, 本文借助注意力机制自适应感知路面病害区域.

图2中, ConvBlock1 和 ConvBlock2 获取原始图像的低维特征表示, 为了在原始图像上获取更大的有效感受野<sup>[30]</sup>, ConvBlock1 采用  $13 \times 13$  的卷积核, 设置卷积步长为 2, 将输入的  $2148 \times 3692$  图像调整为  $1069 \times 1841$ , ConvBlock2 则进一步提高通道数至 32. 之后利用多个改进残差块(residual block)获取更为丰富的特征表示的同时进一步降低特征的长度和宽度. 基于 CBAM 对图像特征进行聚焦, 然后通过 ConvBlock3 和 ConvBlock4 来降低通道数, 还原采样图像, 从而在降低图像分辨率的同时感知保留病害区域相关信息.

### 2.2.1 改进残差模块

从集成学习角度分析, 残差网络可看作是一系列路径集合组装而成的集成模型, 不同的路径包含

嵌入到现有任意 Backbone 之前.

### 2.2 APM 设计

APM 有 2 个功能: 一是降低输入数据的分辨率; 二是自适应感知病害区域相关信息.

了不同的网络层子集<sup>[35]</sup>. 本文借助残差组合低维特征, 获取更为丰富的图像特征表示. 残差模块的一般结构如图3(a)所示, 为了同时实现下采样功能, 本文在图3(a)的基础上增加下采样操作, 如图3(b)所示.

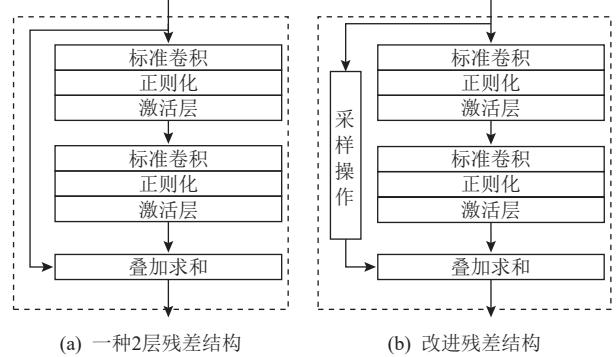


Fig. 3 Residual module

图3 残差模块

APM 可以添加多个残差块, 每个残差块将输入图像变为原来的  $1/2$ , 本文取  $n=2$ , 通过 2 层改进残差结构将特征维度从  $32 \times 1069 \times 1841$  调整为  $32 \times 267 \times 460$ . 残差结构的计算过程表示为式(2):

$$\text{Res}(F_{\text{in}}) = \text{BN}(\text{Conv}_{3 \times 3}(\delta(\text{BN}(\text{Conv}_{5 \times 5}(F_{\text{in}})))) + \text{Sample}(F_{\text{in}})), \quad (2)$$

其中  $F_{\text{in}}$  表示残差块的输入,  $\text{Conv}$  表示卷积操作, 下标表示卷积核尺寸,  $\text{BN}$  表示批规范化(batch normalization),  $\delta$  表示 LeakyReLU 激活函数,  $\text{Sample}(F_{\text{in}})$  则表

示对输入进行下采样操作, 本文中 *Sample* 选择双立方插值算法。

### 2.2.2 CBAM 模块

CBAM 模块包含通道注意力和空间注意力 2 个子模块, 具体结构如图 4 所示。通道注意力模块的目的是激活更有用的通道特征。考虑到不同路面病害的边缘特征不同, 且提取的特征图中的每个通道代表一种特征模式, 把卷积核看作模式检测器, 不同卷积核能够检测出不同的特征模式, 筛选出对边缘信息敏感的卷积核, 即可提取对应的通道特征。空间注意力模块对病害相关区域的空间位置进行感知聚焦, 专注于病害相关的空间位置。

通道注意力对输入  $F_{in}$  进行最大池化和平均池化分别保留突出前景及整体信息, 得到 2 个维度为  $1 \times 1 \times 32$  的通道描述矩阵, 将矩阵输入到共享的 2 层感知机(multi-layer perception, MLP)中, MLP 的第 1 层神经元个数为 4, 激活函数为 *ReLU*, 第 2 层神经元个数为 32。将得到的 2 个特征相加后经过 sigmoid 激活函数得到最终的通道权重系数  $M_c$ ,  $M_c(F_{in}) \times F_{in}$  表示经过通道注意力机制处理后的特征。计算过程由式(3)表示:

$$\begin{cases} M_c(F_{in}) = \delta[MLP(\text{AvgPool}(F_{in})) + \\ \quad MLP(\text{MaxPool}(F_{in}))], \\ F_c = M_c(F_{in}) \times F_{in}. \end{cases} \quad (3)$$

其中,  $\delta$  为 sigmoid 激活函数, MLP 表示多层感知机, *AvgPool* 和 *MaxPool* 分别表示平均池化和最大池化。

空间注意力关注病害区域的空间位置。与通道注意力相似, 输入特征  $F_{in}$  的维度为  $32 \times 267 \times 460$ , 分别进行通道维度的最大池化和平均池化, 得到 2 个维度为  $1 \times 267 \times 460$  的空间位置描述矩阵。将 2 个空间描述矩阵拼接, 经过  $7 \times 7$  的卷积和 sigmoid 激活, 得到维度为  $1 \times 267 \times 460$  的空间权重系数  $M_s$ ,  $M_s(F_{in}) \times F_{in}$  表示

经过空间注意力机制处理后的特征。计算过程由式(4)表示:

$$\begin{cases} M_s(F_{in}) = \\ \quad \delta(\text{Conv}_{7 \times 7}([\text{AvgPool}(F_{in}), \text{MaxPool}(F_{in})])), \\ F_s = M_s(F_{in}) \times F_{in}. \end{cases} \quad (4)$$

### 2.3 损失函数

为了实现端到端的模型训练, 采用联合学习的方式对 APM 进行训练, 考虑本文模型的最终目标是得到最优的路面病害分类结果, 无需额外针对 APM 设计任何损失函数或添加正则化约束, 只需整体利用分类损失进行反向传播即可实现对于 APM 的训练。本文后续实验均采用联合学习的方式完成对 APM 的训练。

本文采用多分类任务中最为常用的交叉熵损失函数, 具体计算过程分 2 步: 第 1 步对网络输出进行 softmax 操作得到对于每个类别的预测概率值, 表示为式(5), 其中  $y_{out}^j$  表示网络的预测输出,  $K$  为网络检测的病害类别总数(本文中  $K=3$ ), softmax 操作将网络的输入映射到一个概率分布维度; 第 2 步对第 1 步的结果取负对数似然即为最终的交叉熵损失函数, 其中  $y_{true}$  为 0-1 编码的真实标签, 如式(6)所示。

$$\text{softmax}(y_{out}^j) = \frac{\exp(y_{out}^j)}{\sum_{k=1}^K \exp(y_{out}^k)}, \quad (5)$$

$$\text{loss} = -\sum_{k=1}^K y_{true}^k \text{lb}(\text{softmax}(y_{out}^k)). \quad (6)$$

### 2.4 APM 可解释性分析

通过可视化对 APM 决策过程的作用机制进行可解释性分析, 度量 APM 对于图像内容的感知区域和感知强度, 从而验证 APM 的有效性。

APM 作为 CNN 技术的一种应用, 目前存在多种基于可视化的可解释性方法来洞察 CNN 的内部卷积

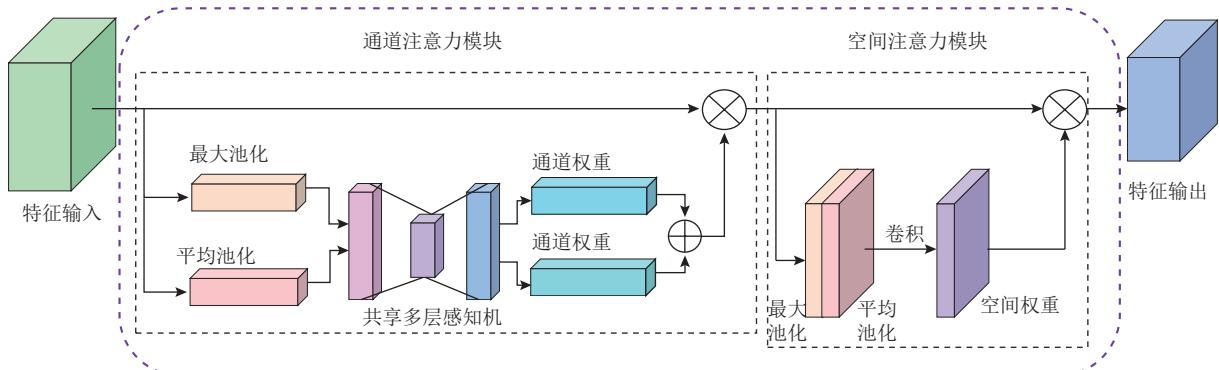


Fig. 4 Convolutional block attention module

图 4 卷积注意力模块

计算过程。从可视化的角度来看，目前主要存在3种可视化模式：特征可视化、卷积核参数可视化、类激活图映射可视化。相比于其他2种可视化技术，类激活图映射可视化通过热力图直观地了解到图像中哪些部分对模型最终的输出结果起决定性作用，同时也可以粗略地定位出图像中物体的位置。在众多类激活图映射可视化范式的方法中，Grad-CAM无需修改模型结构和重新训练模型，适用于不同任务以及多种结构的CNN模型，因此，本文选择Grad-CAM对APM进行可解释性分析和可视化。

要得到APM的对图像内容感知后的输出效果图，只需要计算APM中每层输出的特征图以及对应权重，最后将所有的特征图加权求和后叠加在输入图像上得到APM的类激活图。对于特征图直接取APM中每层卷积的输出即可，对于特征图的权重需要通过反向传播计算目标层所输出的各个特征图对于预测值的梯度，将梯度全局平均池化后得到特征图的权重，整个计算过程可以表示为式(7)：

$$\alpha_k^c = \frac{1}{Z} \sum_{i=1}^w \sum_{j=1}^h \frac{\partial y^c}{\partial A_{ij}^k}, \quad (7)$$

其中 $\alpha_k^c$ 表示类别 $c$ 在第 $k$ 个特征图的对应权重， $Z$ 代表整个特征图的大小， $w, h$ 分别代表特征图的宽度和高度， $y^c$ 表示未经过softmax之前时类别 $c$ 的得分， $A_{ij}^k$ 表示第 $k$ 个特征图中位置 $(i, j)$ 的激活值。

然后，将APM中的所有特征图进行加权求和，并通过ReLU激活函数保证输出激活图中所有数值非负，从而消除与类 $c$ 无关的干扰，得到最终的梯度加权类激活图，计算公式为式(8)。将得到梯度加权类激活图与原始输入图像进行叠加，获得APM的类激活图。

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left( \sum_{k=1}^K \alpha_k^c A^k \right), \quad (8)$$

其中， $K$ 表示特征图的个数。

同样，可以利用Grad-CAM获取骨干网络的类激活图来辅助分析选择最优的骨干网络。

### 3 实验结果与分析

#### 3.1 数据集

本文使用来自Liu等人<sup>[18]</sup>提出的CQU-BPMDD路面病害数据集，该数据集包含11 819张分辨率为 $2148 \times 3692$ 的中国南方地区高速路面的病害图像，包含横向裂纹、纵向裂纹以及修补3种病害类别，随机选择

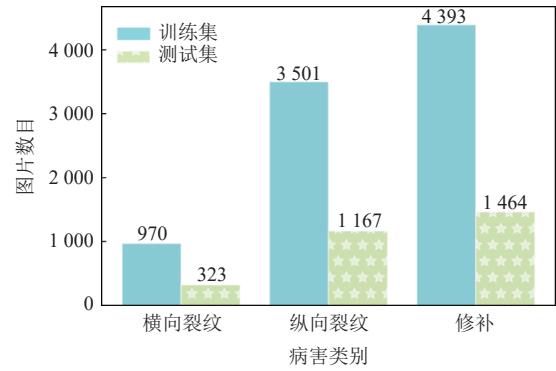


Fig. 5 Data distribution of CQU-BPMDD dataset

图5 CQU-BPMDD数据集数据分布

其中90%的数据作为训练集，10%的数据作为测试集，具体的数据集分布和划分如图5所示。

#### 3.2 实验超参数设置

APM基于开源深度学习框架Pytorch实现，采用在ImageNet上预训练权重的Backbone，选择组合优化器RangerLars，组合了RAdam<sup>[36]</sup>，LookAhead<sup>[37]</sup>，LARS<sup>[38]</sup>，训练时学习率为0.005，批训练样本个数为16，最大训练迭代次数为100，APM最后一层的动态池化的池化大小依据Backbone进行调整。

#### 3.3 评估指标

路面病害分类任务中，可选择的评估指标有精确率(Precision)、召回率(Recall)和F1值。本文选择F1值作为评估指标。F1值的计算借助于精确率和召回率。精确率关注模型在预测结果中为真的准确率，表示在结果为真的时候的模型的可信度，计算公式为式(9)。召回率衡量模型对实际正例的判别能力，表示在当前类别的所有预测结果为正例中真实正例所占比重，计算公式为式(10)。式(9)(10)中，TP表示被正确划分为正例的个数；FP表示被错误划分为正例的个数；FN表示被错误划分为负例的个数；TN表示被正确划分为负例的个数。F1值是对精确率和召回率的一个调和平均，计算公式为式(11)。

$$\text{Precision} = \frac{TP}{TP+FP}, \quad (9)$$

$$\text{Recall} = \frac{TP}{TP+FN}, \quad (10)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (11)$$

采用准确率(Acc)来评估模型的整体性能，计算公式如式(12)所示：

$$Acc = \frac{TP+TN}{TP+FP+TN+FN}. \quad (12)$$

#### 3.4 APM的设计与决策分析实验

为了能够设计出有效的APM模块，首先利用可

视化解释方法来分析不同的结构模型的感知激活效果, 验证设计思路有效性的同时辅助模型的选择和设计。结合可视化的模型决策效果和最终整体路面病害分类的准确率来选择最优的网络结构。

选择 Efficientnetv2\_B2 作为 Backbone, 利用 Grad-CAM 来计算获取激活图, 依次尝试图 6 中的 APM\_Conv, APM\_Res, APM\_CBAM 来分析大卷积核、下采样残差、卷积注意力机制的感知激活效果。借助于联合学习的方式实现端到端的模型训练。

获取 APM\_Conv, APM\_Res, APM\_CBAM, APM

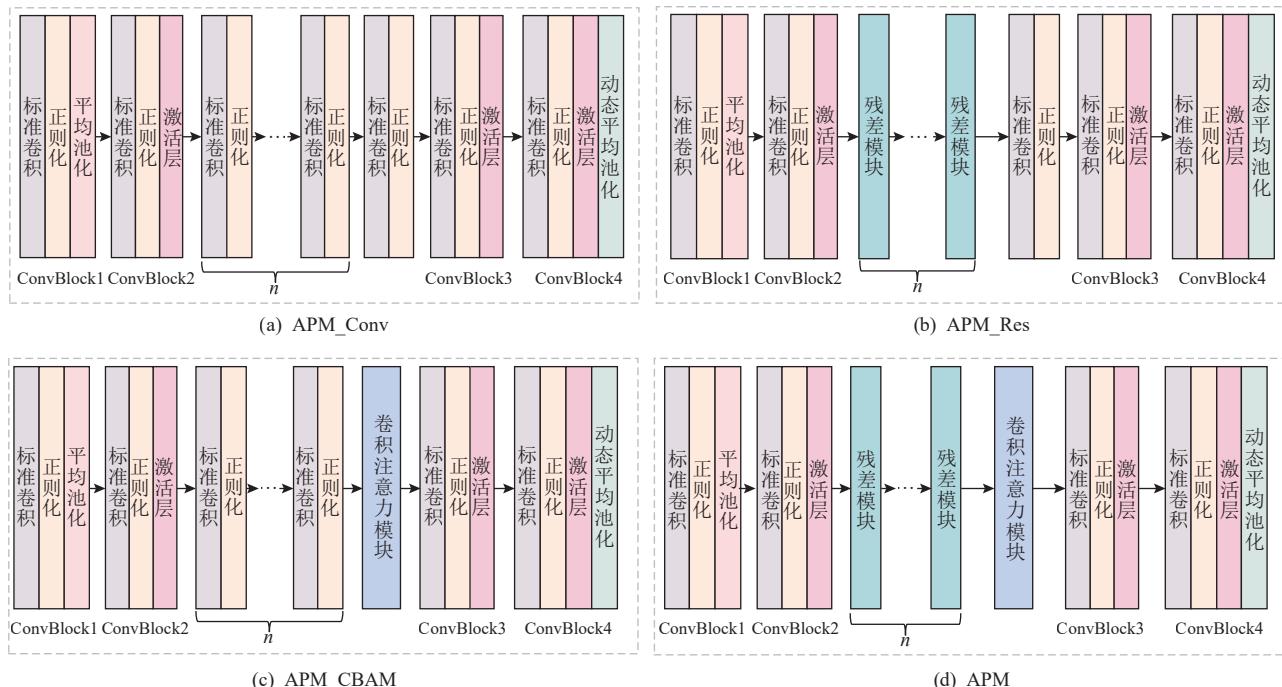


Fig. 6 The detail structures of APM\_Conv, APM\_Res, APM\_CBAM, APM

图 6 APM\_Conv, APM\_Res, APM\_CBAM, APM 的详细结构

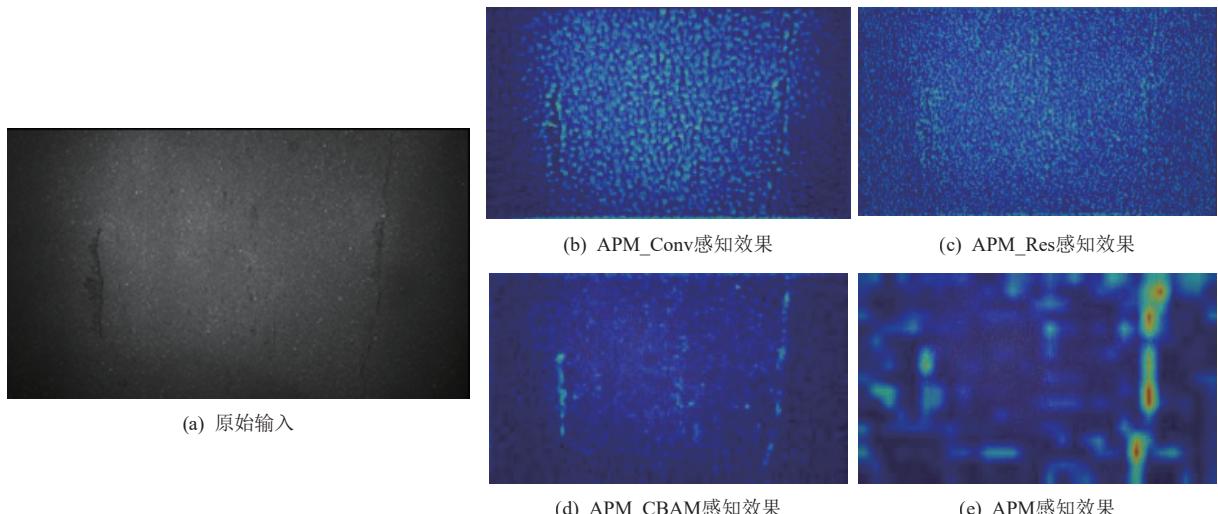


Fig. 7 Adaptive perception of pavement distress region of APM\_Conv, APM\_Res, APM\_CBAM, APM

图 7 APM\_Conv, APM\_Res, APM\_CBAM, APM 的路面病害区域的自适应感知

的感知效果如图 7 所示。图 7(a)为原始高分辨路面输入图像;图 7(b)表示 APM\_Conv 的感知激活效果, 可以看出 APM\_Conv 此时是对图像整体内容的弱感知激活, 说明大卷积核能够有效扩大感受野, 但是并不能有效地感知激活前景病害区域。APM\_Res 的激活效果如 7(c)所示, 残差结构也只是获取更为丰富的特征表示, 也无法有效地感知激活前景病害区域。加入 CBAM 模块后, APM\_CBAM 对路面图像的感知效果如图 7(d)所示, 可以看出路面病害区域在一定程度上被感知激活, 与无关背景信息存在较为明显

的差距，也进一步说明了注意力机制可以有效感知图片中的前景信息。APM 感知效果如图 7(e)所示，可以看到前景病害区域被明显的感知激活，在这 4 种不同结构的 APM 模块中，APM 感知激活效果最好。

利用可视化解释方法辅助设计和选择出内容自适应感知效果明显的 APM，并且其内部决策过程具备良好的可解释性。大卷积核卷积和下采样残差结构可以快速降低输入图像尺寸的同时，获取丰富的

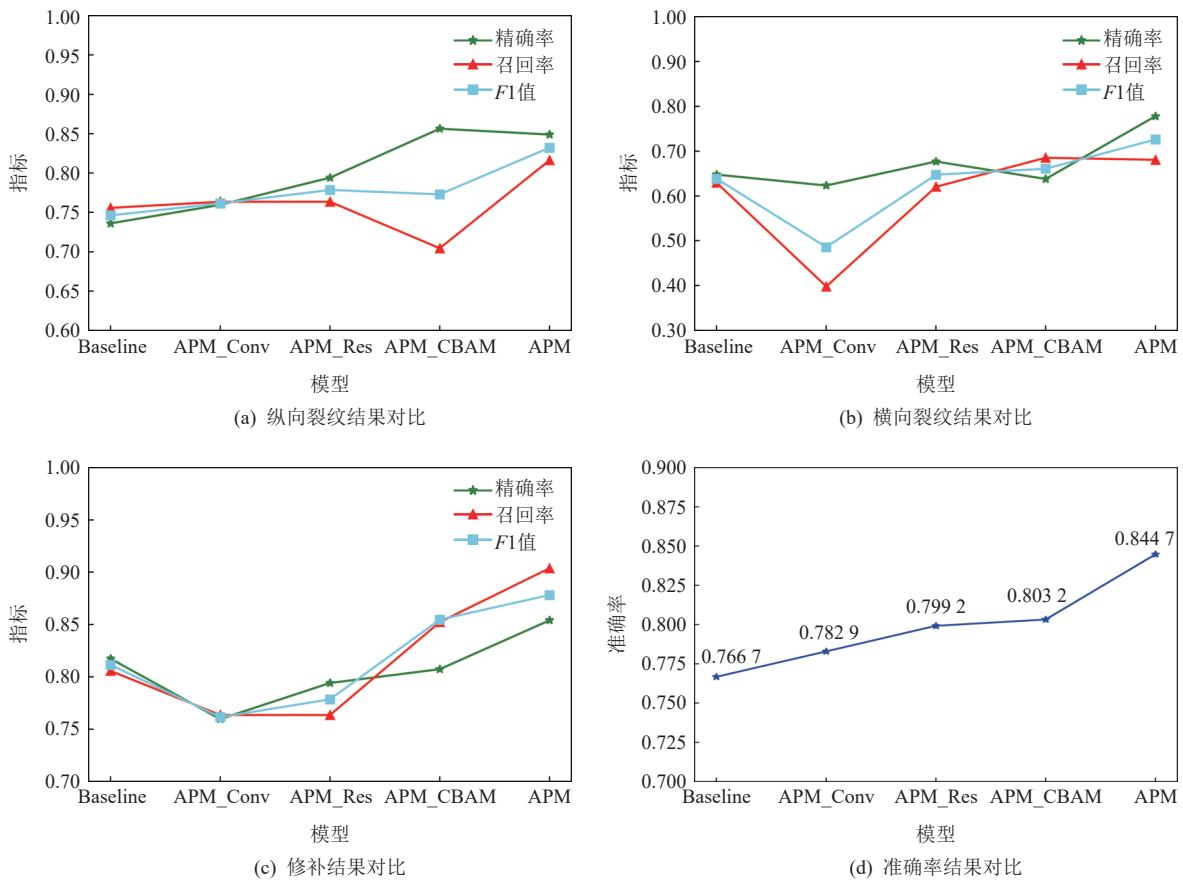


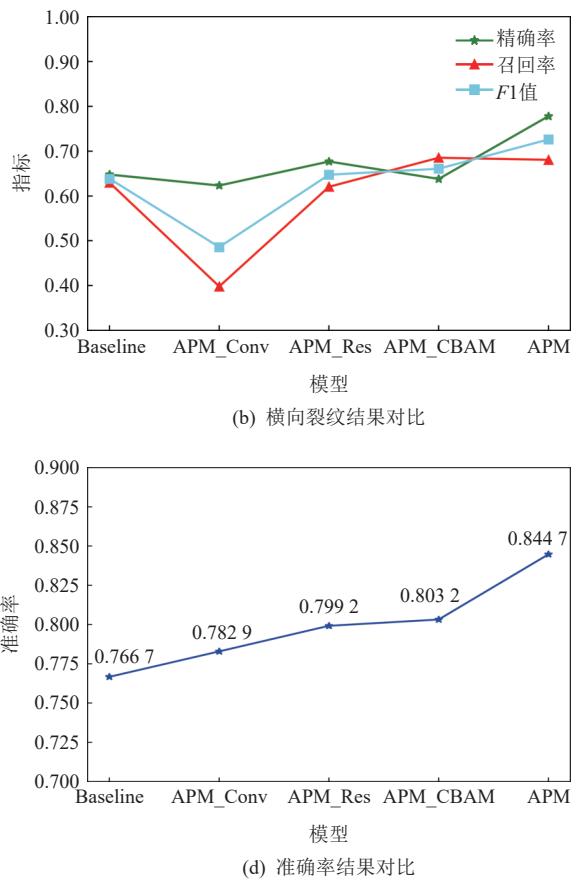
Fig. 8 Comparison of the results of APM\_Conv, APM\_Res, APM\_CBAM, APM

图 8 APM\_Conv, APM\_Res, APM\_CBAM, APM 结果对比

从图 8 可以看出，APM\_Conv 仅利用卷积操作缩放图像，在当前任务上相比于 Baseline 准确率提升约 2.5 个百分点，从整体网络结构来看相当于增加了网络深度，有一定的效果提升。APM\_Res 利用残差模块组合低维特征，准确率相比于 Baseline 提升约 2 个百分点，表明利用残差组合低维特征虽可以产生更为丰富的特征表示，但组合特征同样也包括无关背景特征，提升效果与 APM\_Conv 相近。APM\_CBAM 通过卷积注意力聚焦路面病害区域相关的通道特征和空间特征，准确率相比于 Baseline 提升约 2.8 个百分点，浅层特征无法充分表征原始路面图像语音信息，

语义表示；注意力机制能够有效地自适应感知图像中前景病害区域。

可视化解释方法可以为模型的设计提供直观视觉效果分析，路面病害检测准确率可以定量评估模型的实际效果。在 CQU-BPMDD 数据集下 APM\_Conv, APM\_Res, APM\_CBAM, APM 的分类检测结果对比如图 8 所示，其中 Baseline 采用传统双线性插值下采样算法。



导致 CBAM 不能有效聚焦路面病害区域特征，提升效果与 APM\_Conv, APM\_Res 相当。APM 整体准确率最高达 0.8447，比 Baseline 准确率高约 8 个百分点，对于纵向裂纹、横向裂纹以及修补 3 种病害类别的 F1 值分别提升约 9 个百分点、9 个百分点和 6 个百分点，说明本文对 APM 设计思路的正确性。

结合可视化解释方法对 APM 决策过程直观的分析效果以及在分类检测中准确率等定量评估指标上的有效提升，选择 APM 作为最合理有效的结构。

### 3.5 实验结果及分析

下采样算法分别选择双线性插值(bilinear)、双

立方插值(bicubic)和 Lanczos 插值 3 种传统方法以及 Talebi 等人<sup>[15]</sup>提出的可学习 Resizer 作为对比;选择 ResNet<sup>[39]</sup>, Vit<sup>[40]</sup>, RepVGG<sup>[41]</sup>, Efficientnetv2\_B2<sup>[42]</sup> 分别作为 Backbone 进行实验,对于 APM 和 Resizer 方法,均采用同样的联合学习策略显现端到端的模型训练,实验结果如表 1 所示。

从表 1 可以得出,传统的采样预处理算法在采样过程中同等对待无关背景区域和病害相关区域,缩放图像分辨后,病害区域相关信息的占比并没有提

升,甚至有所降低,导致检测准确率在所有采样预处理算法中最低。Talebi 等人<sup>[15]</sup>提出的 Resizer 倾向于尽可能保留整体的细节,缺乏对路面病害相关的信息的聚焦,保留更多的整体细节信息虽在一定程度上有助于保留更多的模式信息,但同时也保留了无关背景的细节信息,因此相比于传统采样预处理算法准确率虽有一定提升,也仅只有 2 个百分点左右。本文提出 APM 利用注意力机制自适应感知保留病害区域相关信息,在降低原始的图像尺寸的同时可

**Table 1 Comparison of Experimental Results of Different Preprocessing Algorithms and Models**  
**表 1 不同处理算法与模型实验结果对比**

采样算法	Backbone	纵向裂纹			横向裂纹			修补			准确率
		精确率	召回率	F1 值	精确率	召回率	F1 值	精确率	召回率	F1 值	
Bilinear		0.719 3	0.678 6	0.698 1	0.467 8	<b>0.472 2</b>	0.470 0	0.794 1	0.828 2	0.810 8	0.730 2
Bicubic		0.700 2	0.696 6	0.698 4	0.505 2	0.444 4	<b>0.472 9</b>	0.797 6	0.822 0	0.809 6	0.731 2
Lanczos	ResNet	0.715 0	0.696 6	0.705 7	0.480 4	0.453 7	0.466 7	0.800 0	0.826 1	0.812 8	0.734 2
Resizer		0.762 4	0.668 3	0.712 3	0.515 2	0.379 6	0.436 1	0.784 0	<b>0.905 9</b>	0.712 3	0.754 5
APM(本文)		<b>0.767 9</b>	<b>0.714 6</b>	<b>0.740 3</b>	<b>0.524 3</b>	0.389 1	0.452 6	<b>0.815 4</b>	0.903 8	<b>0.857 4</b>	<b>0.773 8</b>
Bilinear		0.743 0	0.758 3	0.750 6	0.644 8	0.638 8	0.641 8	0.815 3	0.803 6	0.809 4	0.767 7
Bicubic		0.754 5	0.750 6	0.752 5	0.603 4	0.648 1	0.625 0	0.819 8	0.809 8	0.814 8	0.768 7
Lanczos	Vit	0.763 4	0.730 0	0.756 3	0.636 4	0.648 1	0.642 2	0.803 6	0.828 2	0.815 7	0.769 7
Resizer		0.788 6	0.748 0	0.767 8	0.641 6	0.712 9	0.675 4	0.824 5	0.838 4	0.831 6	0.789 0
APM(本文)		<b>0.794 2</b>	<b>0.784 0</b>	<b>0.789 1</b>	<b>0.650 0</b>	<b>0.722 2</b>	<b>0.684 2</b>	<b>0.846 4</b>	<b>0.834 3</b>	<b>0.840 3</b>	<b>0.802 2</b>
Bilinear		0.728 1	0.771 2	0.749 0	0.646 4	0.592 5	0.618 3	0.823 1	0.799 5	0.811 2	0.765 7
Bicubic		0.793 1	0.742 9	0.741 0	0.653 4	0.611 1	0.631 6	0.811 7	0.820 0	0.815 8	0.766 7
Lanczos	RepVGG	0.753 2	0.745 5	0.749 3	0.623 9	<b>0.675 9</b>	0.648 9	0.818 2	0.809 8	0.813 4	0.769 7
Resizer		0.780 3	<b>0.776 3</b>	<b>0.778 3</b>	<b>0.660 5</b>	0.666 7	<b>0.663 4</b>	<b>0.828 5</b>	0.830 2	0.829 4	0.791 0
APM(本文)		<b>0.856 2</b>	0.704 3	0.772 9	0.637 9	0.685 1	0.660 7	0.807 2	<b>0.907 9</b>	<b>0.854 5</b>	<b>0.803 2</b>
Bilinear		0.736 8	0.755 7	0.746 1	0.647 6	0.629 6	0.638 5	0.817 4	0.805 7	0.811 5	0.766 7
Bicubic		0.753 9	0.740 3	0.747 1	0.638 1	0.620 3	0.629 1	0.809 6	0.826 1	0.817 8	0.769 7
Lanczos	Efficientnetv2_B2	0.769 0	0.727 5	0.747 7	0.614 0	0.648 1	0.630 3	0.807 5	0.832 3	0.819 7	0.770 7
Resizer		0.793 0	0.758 3	0.775 3	0.622 8	0.657 4	0.639 6	0.826 0	0.844 5	0.835 1	0.790 6
APM(本文)		<b>0.848 9</b>	<b>0.816 2</b>	<b>0.832 2</b>	<b>0.777 8</b>	<b>0.680 5</b>	<b>0.725 9</b>	<b>0.854 0</b>	<b>0.903 8</b>	<b>0.878 2</b>	<b>0.844 7</b>

注: 黑体数值表示最优结果。

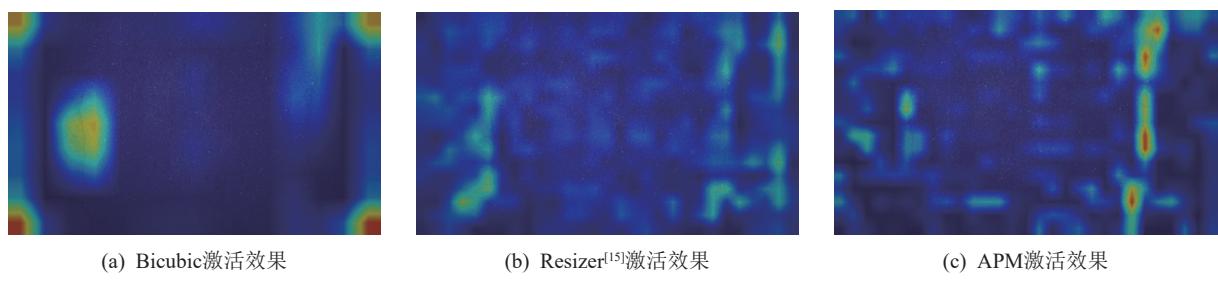


Fig. 9 Adaptive perception of pavement distress region by APM module

图 9 APM 模块对路面病害区域的自适应感知

以感知到与任务紧密相关的路面病害相关区域信息，相比传统下采样算法准确率提升约 8 个百分点，相比于 Talebi 等人<sup>[15]</sup>提出的 Resizer 准确率最高提升约 5 个百分点。

除了分类检测的实验结果定量分析，同样可以借助于可视化的方式来分析对比传统下采样算法、Resizer<sup>[15]</sup>以及 APM 的感知激活效果，具体效果如图 9 所示。当采用传统的图像缩放算法比如双立方插值时，激活效果如图 9(a)所示，可以看出，传统的下采样算法缩放图像后，几乎丢失了右侧细小裂纹的全部信息，激活了过多的无关背景信息，这也进一步解释了传统下采样算法分类检测性能最差的原因。Resizer 的性能提升相比于传统的下采样算法有明显的提升，对其作用结果的可视化如图 9(b)所示。图 9(b)相比于本文提出 APM 的感知效果图 9(c)，虽在一定程度上可以感知激活病害区域，但同时激活了更多的无关背景信息，这也恰好说明了无选择地保留全

部细节信息会在当前任务中引入过多无关背景噪声干扰，影响最终的检测效果。

图 10 对比了传统下采样算法在 4 种不同的 Backbone 下的模型复杂度与分类准确率。其中 B2 表示 Efficientnetv2\_B2，B2, Vit, ResNet, RepVGG 均采用传统的双线性差值下采样算法。同时也对比了 Resizer<sup>[15]</sup>和 APM 在采用 B2 作为 Backbone 下的整体复杂度与准确率。在采用传统的下采样方法时，随着 Backbone 复杂度的上升，准确率没有随之线性提升，甚至有所降低，表明在当前路面病害分类检测任务中对于更深层次的图像语义信息的依赖并不明显，过大过深的模型可能存在过拟合问题。Resizer 相比 APM，模型参数量增加了  $0.05 \times 10^6$ ，且计算复杂度 FLOPs(floating point operations)增加了 21.9 GFLOPs，整体准确率反而降低了 7 个百分点左右，这也进一步说明了同等对待背景与前景信息，缺乏对前景病害区域的关注无法有效地提升路面病害分类检测准确率。

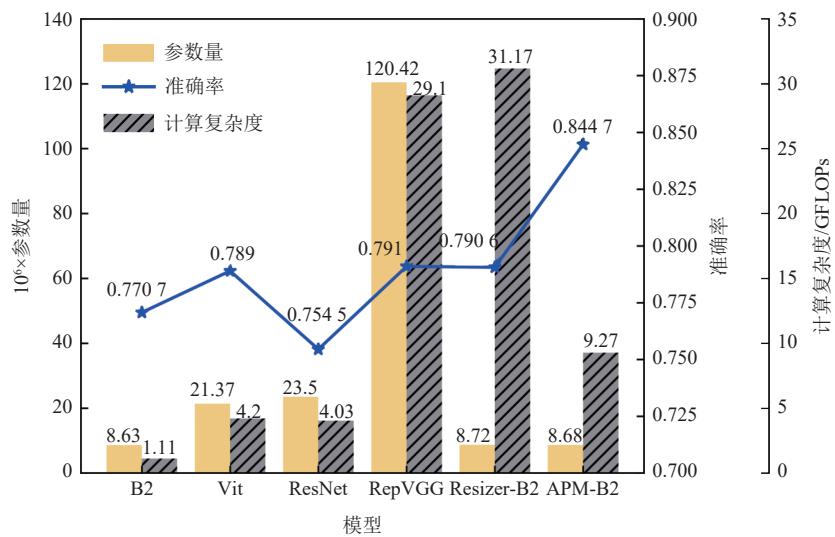


Fig. 10 Comparison between model complexity and accuracy

图 10 模型复杂度和准确率之间的对比

结合表 1 和图 10 可以看出，在最优的 Backbone 下，APM 相比于传统的双线性插值、双立方插值和 Lanczos 插值下采样算法，准确率分别提升 7.8 个百分点、7.5 个百分点、7.4 个百分点；相比于 Talebi 等人<sup>[15]</sup>提出的可学习的 Resizer 准确率提升 5.41 个百分点。此外，相比于传统下采样算法，APM 的模型参数仅增加了  $0.05 \times 10^6$ ，相比于 Resizer 模型参数反而减少了  $0.05 \times 10^6$ 。APM 在仅增加少量模型参数的成本下，整体的病害分类准确率提升显著。

### 3.6 跨数据集验证实验

为了验证本文提出 APM 模块的泛化性与鲁棒

性，本文选择图像分辨率为  $1200 \times 900$  的公开数据 CQU-BPDD<sup>[12]</sup> 进行跨数据集的鲁棒性与泛化性验证实验，CQU-BPDD 共包含了 7 种病害类别，其中训练集共有图片 5140 张，测试集共有图片 11589 张。考虑到 CQU-BPDD 中图像分辨率为  $1200 \times 900$ ，为了避免过度缩放，将 APM 中残差块的数目  $n$  设置为 1，其余设置保持与训练超参数不变，且同样选择联合学习的策略无需额外的人工成本实现对 APM 端到端的模型训练。为了与 CQU-BPDD 数据集的 SOTA(state-of-the-art) 结果进行对比，选择 Top-1 Acc 和对应的 F1 值作为评估指标，实验结果对比如表 2 所示。从表

2可以看出,当使用 Efficientnetv2\_B2 作为 Backbone 时,APM 相比于传统下采样算法在 Top-1 Acc 上提升 2.3 个百分点,但是相比于 Resizer<sup>[15]</sup> 的 Top-1 Acc 提

升不足 1 个百分点。为了探究 APM 效果提升微弱的原因,在使用 APM 时,随机选择样本对模型的类激活图进行可视化展示,如图 11 所示。

Table 2 Comparison of Experimental Results of Different Algorithms on CQU-BPDD Dataset

表 2 在 CQU-BPDD 数据集上不同算法的实验结果对比

采样算法	Backbone	Top-1 Acc	F1 值
Bilinear		0.765 3	0.683 4
Resizer	Efficientnetv2_B2	0.781 6	0.692 3
APM(本文)		<b>0.788 3</b>	<b>0.697 4</b>
Bilinear		0.788 5	0.707 9
Resizer	Efficientnet_B3	0.806 4	0.729 9
APM(本文)		<b>0.846 4</b>	<b>0.810 4</b>
WSPLIN-IP		<b>0.850 0</b>	0.772 0

注: 黑体数值为最优结果。

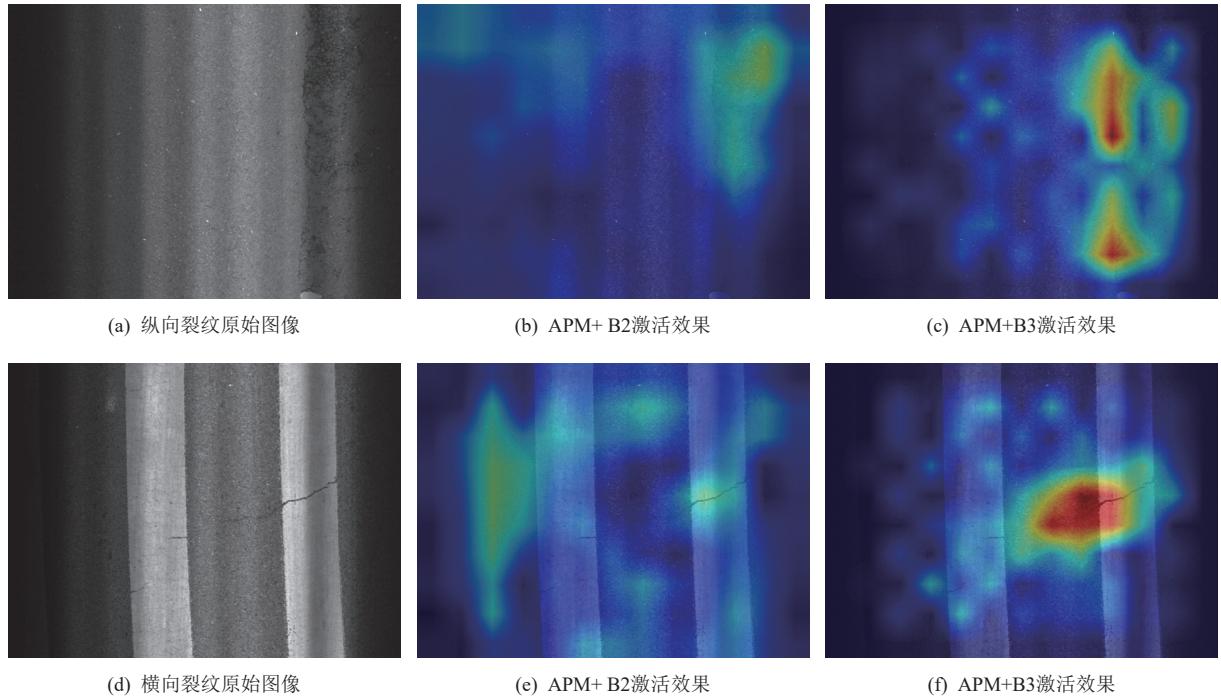


Fig. 11 Comparison of perception of pavement distress region of different Backbone on CQU-BPDD dataset

图 11 CQU-BPDD 数据集上不同 Backbone 对路面病害区域的激活效果对比

图 11(a)(d) 为原路面病害图像,当 Backbone 为 Efficientnetv2\_B2 时,最终的激活效果图如图 11(b)(e) 所示。可以看出激活效果并不明显,对于 CQU-BPDD 数据集,其图像采集环境、光照条件更为复杂,同时存在积水、落叶等干扰因素, Efficientnetv2\_B2 模型较小,容量有限,可能存在欠拟合问题。

考虑 WSPLIN-IP<sup>[14]</sup> 选择 Efficientnet\_B3 作为特征提取器,本文同样选择 Efficientnet\_B3 作为 Backbone。从表 2 可以看出,更换 Backbone 后,APM 的 Top-1 Acc

相比于传统的下采样算法提升约 6 个百分点,相比于 Resizer<sup>[14]</sup> 提升约 4 个百分点。类激活可视化如图 11 所示,可以看出,更换 Backbone 为 Efficientnet\_B3, APM 模块自适应感知区域更为准确和突出。在 CQU-BPDD 数据集上的表现说明了 APM 具有良好的泛化性与鲁棒性。

但是对比于 CQU-BPDD 上的 SOTA 结果,APM 比 WSPLIN-IP<sup>[14]</sup> 低约 0.36 个百分点,WSPLIN-IP<sup>[14]</sup> 采用分块的思想,对分块后的图像进行是否为病害块的

二分类，除了分块策略，WSPLIN-IP<sup>[14]</sup>同时将图像金字塔结构输入到特征提取网络中来获取原始图像的多尺度特征，考虑到病害图像块的数目远小于无病害图像块的数目，利用稀疏采样来加入先验知识。相比于APM，WSPLIN-IP<sup>[14]</sup>额外借助了图像多尺度和先验知识，但是其最终的分类准确率也仅仅只提升了0.36个百分点。

## 4 结 论

本文针对高分辨率图像中病害区域占比低情形下的路面病害检测，借助于可视化解释方法，设计了一种可解释性高分辨图像内容自适应感知模块（APM），用于构建基于视觉的路面病害自动化检测软件系统。具体而言，APM通过大卷积核获取高分辨图像下更大的感受野，利用残差模块组合低维图像特征，借助卷积注意力机制来自适应感知激活路面病害区域相关的特征，从而在降低输入图像尺寸的同时感知激活病害区域相关信息；采用联合学习的方式将APM与现有的骨干网络进行端到端的联合训练，无需额外的监督信息与损失函数，降低APM的训练成本，实现APM即插即用地嵌入到任意Backbone之前。实验证明，本文所提出的APM可以有效感知路面病害区域信息，显著提高了路面病害检测的准确率。APM设计及决策过程的可视化分析实验，说明本文借助可视化解释方法辅助APM设计思路的有效性的同时APM具备良好的解释性。在不同数据集上的实验结果表明，APM具备良好的泛化性和鲁棒性。

**作者贡献声明：**李傲负责实验设计和实现以及论文撰写；葛永新、刘慧君负责实验指导以及论文撰写指导；杨春华负责可视化实验实现；周修庄负责论文撰写指导。

## 参 考 文 献

- [1] Cai Jun, Jiang Lizheng. Cause analysis and maintenance countermeasure of common distresses of expressway asphalt pavement [J]. Engineering and Construction, 2006, 20(1): 81–83 (in Chinese) (蔡骏, 蒋立政. 高速公路沥青路面常见病害的成因分析及养护对策 [J]. 工程与建设, 2006, 20(1): 81–83)
- [2] Liu Fan, Wang Junfeng, Chen Zhiyu, et al. Parallel attention based UNet for crack detection [J]. Journal of Computer Research and Development, 2021, 58(8): 1718–1726 (in Chinese)
- (刘凡, 王君锋, 陈峙宇, 等. 基于并行注意力UNet的裂缝检测方法 [J]. 计算机研究与发展, 2021, 58(8): 1718–1726)
- [3] Luo Hui, Jia Chen, Li Jian. Road surface distress detection algorithm based on improved YOLOv4 [J]. Laser & Optoelectronics Progress, 2021, 58(14): 336–344 (in Chinese) (罗晖, 贾晨, 李健. 基于改进YOLOv4的公路路面病害检测算法 [J]. 激光与光电子学进展, 2021, 58(14): 336–344)
- [4] Wang Ye, Chen Junwu, Xia Xin, et al. Intelligent requirements elicitation and modeling: A literature review [J]. Journal of Computer Research and Development, 2021, 58(4): 683–705 (in Chinese) (汪烨, 陈骏武, 夏鑫, 等. 智能需求获取与建模研究综述 [J]. 计算机研究与发展, 2021, 58(4): 683–705)
- [5] Amhaz R, Chambon S, Idier J, et al. Automatic crack detection on two-dimensional pavement images: An algorithm based on minimal path selection [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(10): 2718–2729
- [6] Koch C, Georgieva K, Kasireddy V, et al. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure [J]. Advanced Engineering Informatics, 2015, 29(2): 196–210
- [7] Kapela R, Śniatała P, Turkot A, et al. Asphalt surfaced pavement cracks detection based on histograms of oriented gradients [C]// Proc of the 22nd Int Conf Mixed Design of Integrated Circuits & Systems. Piscataway, NJ: IEEE, 2015: 579–584
- [8] Ju Zhuoya, Wang Zhihai. A Bayesian classification algorithm based on selective patterns [J]. Journal of Computer Research and Development, 2020, 57(8): 1605–1616 (in Chinese) (鞠卓亚, 王志海. 基于选择性模式的贝叶斯分类算法 [J]. 计算机研究与发展, 2020, 57(8): 1605–1616)
- [9] Fu Zhongliang. Real AdaBoost algorithm for multi-class and imbalanced classification problems [J]. Journal of Computer Research & Development, 2011, 48(12): 2326–2333 (in Chinese) (付忠良. 不平衡多分类问题的连续AdaBoost算法研究 [J]. 计算机研究与发展, 2011, 48(12): 2326–2333)
- [10] Gopalakrishnan K, Khaitan S K, Choudhary A, et al. Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection [J]. Construction and Building Materials, 2017, 157: 322–330
- [11] Li Shengyuan, Zhao Xuefeng. Convolutional neural networks-based crack detection for real concrete surface [C]// Proc of the 23rd Conf on Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems. Bellingham, WA: SPIE, 2018, 10598: 955–961
- [12] Cha Youngjin, Choi W, Büyüköztürk O. Deep learning-based crack damage detection using convo-lutional neural networks [J]. Computer-Aided Civil and Infrastructure Engineering, 2017, 32(5): 361–378
- [13] Tang Wenhao, Huang Sheng, Zhao Qiming, et al. An iteratively optimized patch label inference network for automatic pavement distress detection [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(7): 8652–8661
- [14] Huang Guixin, Huang Sheng, Huangfu Luwen, et al. Weakly supervised patch label inference network with image pyramid for pavement distress recognition in the wild [C]// Proc of the 46th IEEE Int Conf on Acoustics, Speech and Signal Processing. Piscataway, NJ:

- IEEE, 2021: 7978–7982
- [15] Talebi H, Milanfar P. Learning to resize images for computer vision tasks[C]// Proc of the 20th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 497–506
- [16] Wang Ziji, Turko R, Shaikh O, et al. CNN explainer: Learning convolutional neural networks with interactive visualization[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2020, 27(2): 1396–1406
- [17] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]// Proc of the 15th European Conf on Computer Vision. Berlin: Springer, 2018: 3–19
- [18] Liu Huijun, Yang Chunhua, Li Ao, et al. Deep domain adaptation for pavement crack detection[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 24(2): 1669–1681
- [19] Wang Weixing, Wang Mengfei, Li Hongxia, et al. Pavement crack image acquisition methods and crack extraction algorithms: A review[J]. *Journal of Traffic and Transportation Engineering (English Edition)*, 2019, 6(6): 535–556
- [20] Jenkins M D, Carr T A, Iglesias M I, et al. A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks[C]// Proc of the 26th European Signal Processing Conf. Piscataway, NJ: IEEE, 2018: 2120–2124
- [21] Yang Fan, Zhang Lei, Yu Sijia, et al. Feature pyramid and hierarchical boosting network for pavement crack detection[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 21(4): 1525–1535
- [22] Liu Wenjun, Huang Yuchun, Li Ying, et al. FPCNet: Fast pavement crack detection network based on encoder-decoder architecture[J]. arXiv preprint, arXiv:1907.02248, 2019
- [23] Zhang Kaige, Zhang Yingtao, Cheng Hengda. CrackGAN: Pavement crack detection using partially accurate ground truths based on generative adversarial learning[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 22(2): 1306–1319
- [24] Xiang Xuezhi, Zhang Yuqi, El Saddik A. Pavement crack detection network based on pyramid structure and attention mechanism[J]. *IET Image Processing*, 2020, 14(8): 1580–1586
- [25] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]// Proc of the 28th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 3431–3440
- [26] Zhang Lei, Yang Fan, Zhang Y D, et al. Road crack detection using deep convolutional neural network[C]// Proc of the 23rd IEEE Int Conf on Image Processing. Piscataway, NJ: IEEE, 2016: 3708–3712
- [27] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84–90
- [28] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proc of the 27th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 580–587
- [29] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]// Proc of the 13th European Conf on Computer Vision. Berlin: Springer, 2014: 818–833
- [30] Zeiler M D, Krishnan D, Taylor G W, et al. Deconvolutional networks[C]// Proc of the 23rd IEEE Computer Society Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2010: 2528–2535
- [31] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91–110
- [32] Zeiler M D, Taylor G W, Fergus R. Adaptive deconvolutional networks for mid and high level feature learning[C]// Proc of the 13th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2011: 2018–2025
- [33] Zhou Bolei, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]// Proc of the 29th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2921–2929
- [34] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]// Proc of the 15th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 618–626
- [35] Veit A, Wilber M J, Belongie S. Residual networks behave like ensembles of relatively shallow networks [J/OL]. *Advances in Neural Information Processing Systems*, 2016 [2023-03-18]. <https://proceedings.neurips.cc/paper/2016/hash/37bc2f75bf1bcfe8450a1a41c200364c-Abstract.html>
- [36] Liu Liyuan, Jiang Haoming, He Pengcheng, et al. On the variance of the adaptive learning rate and beyond[J]. arXiv preprint, arXiv: 1908.03265, 2019
- [37] Zhang M, Lucas J, Ba J, et al. Lookahead optimizer:  $k$  steps forward, 1 step back[J/OL]. *Advances in Neural Information Processing Systems*, 2019 [2023-03-18]. <https://proceedings.neurips.cc/paper/2019/hash/90fd4f88f588ae64038134f1eeaa023f-Abstract.html>
- [38] You Yangs, Gitman I, Ginsburg B. Large batch training of convolutional networks[J]. arXiv preprint, arXiv: 1708.03888, 2017
- [39] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]// Proc of the 29th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770–778
- [40] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth  $16 \times 16$  words: Transformers for image recognition at scale[J]. arXiv preprint, arXiv: 2010.11929, 2020
- [41] Ding Xiaohan, Zhang Xiangyu, Ma Ningning, et al. RepVGG: Making VGG-style convnets great again[C]// Proc of the 34th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 13733–13742
- [42] Tan M, Le Q. Efficientnet\_v2: Smaller models and faster training[C]// Proc of the 38th Int Conf on Machine Learning. Bellevue, WA: ILMS, 2021: 10096–10106



**Li Ao**, born in 1994. Master. His main research interests include computer vision and deep learning.

李傲, 1994年生. 硕士. 主要研究方向为计算机视觉、深度学习。



**Ge Yongxin**, born in 1981. PhD, professor, PhD supervisor. His main research interests include computer vision, machine learning, and big data intelligence.

葛永新, 1981 年生. 博士, 教授, 博士生导师. 主要研究方向为计算机视觉、机器学习、大数据智能.



**Yang Chunhua**, born in 1998. Master. His main research interests include deep learning and digital image processing.

杨春华, 1998 年生. 硕士. 主要研究方向为深度学习和数字图像处理.



**Liu Huijun**, born in 1975. PhD, associate professor, master supervisor. His main research interests include data mining, big data intelligence, and computer vision.

刘慧君, 1975 年生. 博士, 副教授, 硕士生导师. 主要研究方向为数据挖掘、大数据智能、计算机视觉.



**Zhou Xiuzhuang**, born in 1974. PhD, professor, PhD supervisor. His main research interests include computer vision, machine learning, and medical AI.

周修庄, 1974 年生. 博士, 教授, 博士生导师. 主要研究方向为计算机视觉、机器学习、医疗 AI.